Getting Technical with TV&C...

Volume 2, Issue 2 Fall 2000

Item Analysis Data

Introduction

Item analysis data provides some important information regarding the quality of a written examination that has been administered to a group of examinees. As such, the review and interpretation of the data offers an indication of the degree of confidence that can be placed in the exam results. This monograph will provide an introduction and overview of the basic principles underlying the interpretation of item analysis data.

When to Obtain Item Analysis Data

Item analysis data should be obtained following the administration of the exam but prior to finalizing exam scores, setting a pass point, and notifying examinees of results. An item analysis provides useful information about each individual item within the exam, as well as summary information on each exam segment. This information is needed to ensure that the exam is performing as intended and is providing meaningful and interpretable results. Although some basic item analysis data can be calculated by hand, the most detailed and comprehensive data is generated by item analysis software programs.

What Item Analysis Data Is Used For

The objective of an individual exam item used within the employment testing discipline is to discriminate between the better candidates and poorer candidates with respect to a specific criterion. When an exam item is functioning properly, candidates with a knowledge of the target subject matter will answer the item correctly, while candidates who lack the necessary knowledge will fail to answer the item correctly. Those exam items that discriminate effectively between the better and the poorer candidates provide the greatest utility to the exam.

The item analysis data is used to identify the exam items that simply do not discriminate between the better candidates and the poorer candidates. These items are typically problematic and contribute little to the overall effectiveness of the exam. The item analysis data also provides information which is used to identify miskeyed or malfunctioning items, as well as items that may be in need of revision prior to future administrations of the exam. The remainder of this monograph will provide a more detailed discussion of the item analysis data.

Item Statistics

The item analysis provides a significant amount of detailed information about each individual item within the exam. As an aid in evaluating the effectiveness of each exam item, the item analysis splits the candidates into three groups based upon their scores on the particular exam segment. To elaborate, candidate scores on the exam segment are arrayed in descending order so that the 27% of the candidates with the highest scores on the segment are placed in the upper group, the 27% of the candidates with the lowest scores on the segment are placed in he lower group, and the remaining 46% of the candidates are placed in the middle group. From this distribution of scores, the item analysis provides a frequency count, as well as the corresponding percentage value, of the candidates within these upper, middle, and lower groups who selected the correct answer (i.e., the keyed response) to a given item. Similar information is also provided for each of the item distracters. Ideally, all of the candidates in the upper group will answer the item correctly, while all of the candidates in the lower group will get the item wrong. The total percentage of candidates selecting the correct answer is used as an index of the item's difficulty. An item difficulty index between .40 and .60 is desired. That is, ideal items are items that approximately 40 to 60 percent of the candidates answer correctly.

Discrimination Index. In addition to looking at how the candidates in the upper 27% group performed on a given item in comparison to the candidates in the lower 27% group, another way to assess the discriminability of the item is to look at the item discrimination index. The item analysis provides an item discrimination index, in the form of an item reliability coefficient, for each item within the exam segment. This coefficient provides a measure of how well an exam item discriminates between the better and the poorer candidates.

The most common item discrimination indices are the point-biserial correlation and the biserial correlation. Each of these indices provides an indication of the *direction* and the *strength* of the relationship between segment score (a continuous variable) and item score (a dichotomous variable). The item discrimination statistic used by the State Personnel Board's on-line exam system is the point-biserial correlation.

Direction. A *positive* point-biserial correlation coefficient for a response option (e.g., alternative A, B, C, or D) indicates that candidates who performed well on the exam segment also selected that particular response option. Thus, the point-biserial correlation coefficient for the keyed response should always be positive. When this is the case, the coefficient indicates that those candidates who performed well on the exam segment selected the correct answer to the item.

A *negative* point-biserial correlation coefficient for a response option indicates that candidates who performed well on the exam segment *did not* choose that particular response option. Consequently, the point-biserial correlation coefficients for an item's distracters should always be negative. When this is indeed the case, candidates who performed well on the exam segment are avoiding the distracters (i.e., answering the item correctly), while those who performed poorly on the segment are selecting the distracters (i.e., answering the item incorrectly). A negative point-biserial correlation coefficient for the keyed response is an indication that the item is problematic. The problem may simply be that the item has been miskeyed, or the item may be

ambiguous, confusing, or malfunctioning for some other reason.

Strength. Strength refers to the size or magnitude of the point-biserial correlation coefficient. Theoretically, a point-biserial correlation coefficient can range in value from -1.00 to +1.00 (although in actuality the range is restricted to approximately -.80 to +.80). A high point-biserial coefficient for the keyed response is desired since this indicates that those candidates who did well on the exam segment answered the item correctly, while those candidates who performed poorly on the segment answered the item incorrectly. Therefore, the item is doing a good job of discriminating between the better and the poorer candidates. A point-biserial correlation coefficient of .30 and above for the keyed response is considered to be good, while point-biserial coefficients of -.30 and below for the distracters are desirable.

Segment Statistics

In addition to the detailed information about each individual item within the exam, the item analysis also provides summary statistics for each exam segment. Segment statistics typically include an estimate of the segment's reliability, as well as the mean score, standard deviation, mean item difficulty, and the number of items included in the segment.

Reliability. Reliability is defined as the extent to which scores achieved on the exam segment are precise or stable indicators of the candidates' true level of knowledge or skill. When reliability is low, there is an increased chance of accepting candidates who cannot perform the job, or eliminating candidates who can perform the job.

Exam segment reliability for the State Personnel's Board's item analysis reports is estimated using Cronbach's Coefficient Alpha. Coefficient Alpha can range in value from 0.0 to +1.0. The greater the coefficient's magnitude, the more reliable the exam segment. Coefficients greater than .80 are desirable.

Mean. The mean indicates the average candidate score in the distribution of scores. The mean is a measure of central tendency. Measures of central tendency are indicators of the distribution's average or typical score.
Standard Deviation. The standard deviation is a measure of the dispersion of the exam scores around the mean score. In effect, the standard deviation is the average of the difference of the candidates' scores from the mean score. The larger the standard deviation, the more the scores differ from each other. A relatively large standard deviation is desirable for an exam segment.

Mean Item Difficulty. Mean item difficulty is the average item difficulty for the exam segment. A mean item difficulty between .40 and .60 is desirable. Mean item difficulty values of this magnitude indicate that, on average, 40 to 60 percent of the candidates are correctly answering a given item.

Number of Items. Number of items refers to the total number of exam questions within the particular segment. In order to achieve acceptable exam segment reliabilities, each segment should consist of approximately 30 or more items.

Conclusion

An item analysis provides extremely valuable information about the quality of individual exam items as well as summary information on each exam segment. It is important that this item analysis data be reviewed and that necessary remedial action be taken prior to

finalizing exam scores, establishing a pass point, and notifying candidates of their exam results. Remedial action may involve merely correcting the keyed response of a miskeyed item, or deleting malfunctioning items before the exam results are finalized. Item analysis data is also useful for identifying ineffective or obviously incorrect distracters so that they can be revised for future use. In any event, the interpretation and use of the item analysis data is a critical component of the exam development and administration process.

This monograph is intended to provide a general introduction to the basic principles related to item analysis data. A more complete discussion of this topic can be found in the State Personnel Boards publication entitled *Interpreting Item Analysis Data*, and in most textbooks on psychometric theory. The State Personnel Board also offers a one-day training class on item analysis.

Published by the Test Validation and Construction Unit
of the California State Personnel Board.
Visit our Webpage at www.spb.ca.gov/tvchome.htm.